D374–D379 *Nucleic Acids Research, 2014, Vol. 42, Database issue* *Published online 29 September 2013*
doi:10.1093/nar/gkt887

3did: a catalog of domain-based interactions of known three-dimensional structure

Roberto Mosca[1], Arnaud Céol[1,2], Amelie Stein[3], Roger Olivella[1] and Patrick Aloy[1,4,*]

[1]Joint IRB-BSC Program in Computational Biology, Institute for Research in Biomedicine (IRB Barcelona), c/ Baldiri Reixac 10-12, 08028 Barcelona, Spain, [2]Center for Genomic Science of IIT@SEMM, Istituto Italiano di Tecnologia (IIT), Via Adamello 16, 20139 Milan, Italy, [3]California Institute for Quantitative Biomedical Research (qb3) and Department of Bioengineering and Therapeutic Sciences, MC 2530, University of California San Francisco (UCSF) CA 94158–2330, USA and [4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, 08010 Barcelona, Spain

Received August 28, 2013; Revised and Accepted September 9, 2013

ABSTRACT

The database of 3D interacting domains (3did, available online for browsing and bulk download at http://3did.irbbarcelona.org) is a catalog of protein–protein interactions for which a high-resolution 3D structure is known. 3did collects and classifies all structural templates of domain–domain interactions in the Protein Data Bank, providing molecular details for such interactions. The current version also includes a pipeline for the discovery and annotation of novel domain–motif interactions. For every interaction, 3did identifies and groups different binding modes by clustering similar interfaces into 'interaction topologies'. By maintaining a constantly updated collection of domain-based structural interaction templates, 3did is a reference source of information for the structural characterization of protein interaction networks. 3did is updated every 6 months.

INTRODUCTION

Proteins are key players in virtually all events that take place within and between cells. However, they seldom act alone and it is their complex interrelationships that will ultimately determine the behavior of a biological system. For this reason, large efforts have been devoted to unveiling the complex network of interactions between proteins underlying biological processes, producing large interactomes for several organisms, including human (1–3). High-throughput interaction discovery experiments provide valuable information as to who-interacts-with-whom but, to fully understand how protein interactions occur, we need to incorporate high-resolution molecular/atomic details, which are currently available in the Protein Data Bank [PDB, (4)].

Several efforts over the last years aimed at mining the data in the PDB to provide a comprehensive structural characterization of protein interaction networks (5–7). While these studies took different approaches they all agree on one point: interactions are often achieved by the reuse of evolutionary conserved structural modules, represented by domain families. Domains can be found in interaction with other domains (domain–domain interactions or DDIs) or with short, usually structurally extended peptides described by a recurring motif of amino acids (domain–motif interactions or DMIs). The possibility of producing a complete and exhaustive mapping of structural data on protein interactomes depends, therefore, on the availability of a reliable and extended catalog of domain-based 3D structural templates. Given the rate at which new interactions are discovered and new structures of complexes are experimentally characterized, it is paramount for this catalog to be constantly updated.

Several bioinformatics studies have attempted to define and classify domain interactions, both DDIs (8–13) and DMIs (14–17), and produced databases of domain-related interaction models but many of them are not regularly updated or are not available anymore. All these databases also vary in the way they define domains. Some of them use the definition provided by SCOP (18) or CATH (19), which are based on the analysis of experimental structures and are known to lag behind the status of PDB by several years.

The database of 3D interacting domains (3did) is a collection of 3D structures of domain-based interactions, both DDIs and DMIs, based on domain definitions from Pfam (20), ensuring a higher coverage of the

*To whom correspondence should be addressed. Tel: +34 934039690; Fax: +34 934039954; Email: patrick.aloy@irbbarcelona.org

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

protein sequences universe. It has been constantly available to the scientific community for more than 8 years (21–23). With the current version it integrates a pipeline for the automatic identification of novel domain–peptide interactions. Periodic updates will be performed every 6 months to reflect the latest contents of the PDB and the latest definitions of domain families from Pfam. All these characteristics make 3did a reference catalog of domain-based interaction and an essential component for the structural characterization of protein interaction networks.

DOMAIN–DOMAIN INTERACTIONS

DDIs occur when two globular domains form a stable interface. Interfaces in DDIs are usually relatively large [2000 $Å^2$ on average (24)]. Several possible definitions of domains are available based on conserved globular structure or on evolutionary conserved residue sequences. For instance, SCOP (18) and CATH (19) are two catalogs of structurally conserved domains. In 3did we use the domains definitions provided by Pfam, generated from representative homologous protein that are searched against large datasets of protein sequences. Pfam domains, being defined on evolutionary conserved modules at the sequence level, have the advantage of showing a higher coverage of the sequence space. Due to the faster collection of protein sequences, Pfam definitions are updated more often than structure-based domain definitions. The current version of 3did uses Pfam version 27.0, which includes more than 14 000 domain families. Domains are searched on the sequences of all chains present in the PDB by using the pfam_scan.pl script provided by Pfam [which uses HMMER3 (25)]. All nonoverlapping hits are retained. In case of pairs of domains where one overlaps with the center (in sequence) of the other, only the domain with the highest score is retained. We exclude chains shorter than 11 residues, chains reporting only the position of Cα atoms and those where only the backbone has been traced.

We estimate the number of residue–residue interactions between pairs of contacting domains either within the same chain (intrachain) or between two different chains (interchain). We require at least five estimated contacts [hydrogen bonds, electrostatic or van der Waals interactions, as described in (26)] in order to account for an interaction between the two domains. Finally we assign a *z*-score to the DDIs [based on (26,27)]. For every pair of interacting domains, we cluster the corresponding structural templates on the basis of the interaction interface in order to characterize different modes of interaction between the same pair of domains, as described previously (23).

The current version of 3did contains 258 079 structural instances of DDIs of which 68 861 are intrachain and 189 218 are interchain. These correspond to 8328 unique domain–domain pairs (1190 with only intrachain instances and 5747 with only interchain instances while 1391 have both intra and interchain instances). With respect to the last version of 3did (2011) we observed an impressive growth of 62% in the number of DDI structures corresponding to 39.5% more domain–domain distinct (i.e. nonredundant) pairs, reflecting the constantly increasing rate of growing of the PDB and Pfam (Figure 1; please note that we have introduced a release numbering scheme based on the year and month of release: the current version is 2013_06). Table 1 reports the top 10 domains ranked on the number of partner domains. The table also shows that the PDB contains highly redundant data for DDIs. In fact, for every pair of interacting domains, usually there are several structural instances of that DDI, showing, in many cases, different interaction topologies and, sometimes, multiple instances for the same topology.

DOMAIN–MOTIF INTERACTIONS

Domains have also been observed to bind short linear motifs, which show considerably smaller interfaces than those in DDIs [350 $Å^2$ on average (24)]. Given the smaller interface, DMIs are often weaker in nature and thus often used in transient associations such as signaling networks (28). Only a small number of key residues are required for binding, allowing fast evolution of these interactions (29). However, the short motifs are harder to detect automatically than evolutionary conserved domain fingerprints, therefore many resources of domain–motif interactions, such as ELM (30), rely on manual curation. Interactome-wide approaches rely on motifs and protein interaction data to suggest DMIs

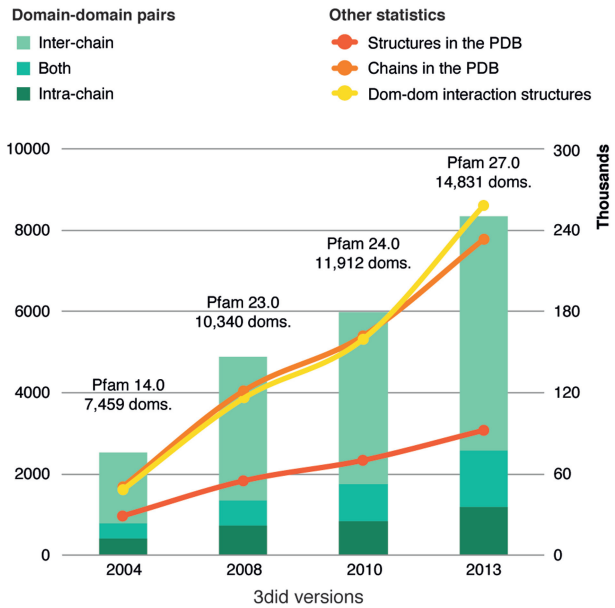

Figure 1. Growth of 3did throughout its four releases. The colored bars represent the number of DDI pairs with only intrachain structural templates (dark green), only interchain templates (medium green) and both types of templates (light green). Bar plots refer to the *y*-axis on the left. The lines represent the growth in the number of structures (dark orange) and chains (light orange) in the PDB. The yellow line represents the number of domain–domain structural templates in 3did (i.e. the number of redundant structural instances of DDI). Line plots refer to the *y*-axis on the right.

Table 1. Top 10 interacting domains with the corresponding number of protein partners. DDI pairs in 3did have variable numbers of structural templates. For example, even if the C1-set domain has less interacting domains than the V-set domain, it has many more redundant structural templates in the PDB

Domain name	Pfam id	# partners	#interaction structures
V-set	PF07686	161	8962
Ras	PF00071	62	610
Pkinase	PF00069	54	1888
Trypsin	PF00089	50	1753
ubiquitin	PF00240	43	632
C1-set	PF07654	39	9114
WD40	PF00400	32	2205
EF-hand_7	PF13499	32	713
Ig_2	PF13895	29	312
Ank_2	PF12796	29	428

[e.g. (31)]. As an alternative approach to DMI detection, commonly observed structural features of these interactions have enabled automated searches in the PDB (32,33). In both interactome- and structure-based approaches, the main challenge is to separate spurious hits from truly over-represented domain–motif pairs. This is usually performed by calculation of statistical significance against a random background as well as enrichment in alternative datasets, such as interactomes of different species. The approach now included in 3did has been described in detail (33) and is outlined in Figure 2. Previous versions of 3did reported only one motif for each DMI topology, even if multiple were found to be significant, while we now report all significant motifs.

The DMI-collection in 3did now contains peptides binding to 113 distinct domains, an increase of ~2.5-fold over the 46 domains described in our 2010 article (33). This goes along with a ~3-fold increase in structures of DMIs, from 1500 to 4500. Since the discovery of novel DMIs requires intensive computation, we have decided to rebuild the contents of the database every 6 months, synchronized with the update of our other structural database Interactome3D [http://interactome3d.irbbarcelona.org, (6)].

3did NEW INTERFACE

The web interface has been entirely redesigned to allow an easier and more enjoyable search. The home page displays basic statistics about the domains and motifs present in the database and informs the user about the versions of Pfam and PDB that are currently used. The results in the database may vary from one version to the other and the user should be aware of each update. The home page also permits a simple query to 3did for a domain or a motif. The different search tools available in the previous version of 3did have been grouped in a single search page. This page allows to search for a domain (either the name of the domain or its Pfam accession number can be used), a motif name, a structure (by PDB ID) or any term from the Gene Ontology (34). The

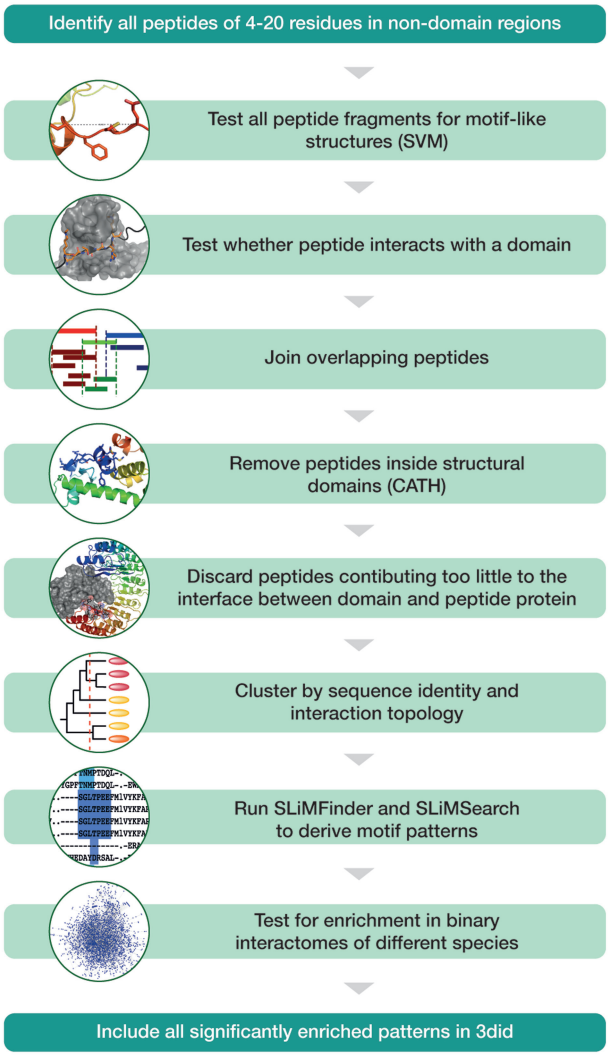

Figure 2. Overview of the DMI discovery pipeline. The main steps of the DMI discovery pipeline are outlined, with filtering steps to remove spurious hits. Details can be found in (33).

association between Pfam accessions and GO terms is downloaded from the Gene Ontology website (http://www.geneontology.org). Alternatively it is possible, through the 'browse' tab, to browse all domains and motifs present in 3did or to explore a GO tree and retrieve all the domains associated to any GO term (Figure 4C).

The data in 3did are displayed in four different views: the domain, motif, interaction and PDB views (Figure 3). The domain view is composed of three parts. The first part shows, both graphically and as a list, the domains and motifs that interact with the query domain. In the graphical interface, based on CytoscapeWeb (35), the interacting domains are displayed in orange and the interacting motifs in green (Figure 4A). A set of four buttons below the graph allows updating the network and displaying the GO terms associated to each domain. The interacting domains and motifs are also displayed as a list. Both lists and the network are linked to the page describing the domain or motif and the corresponding DDIs or

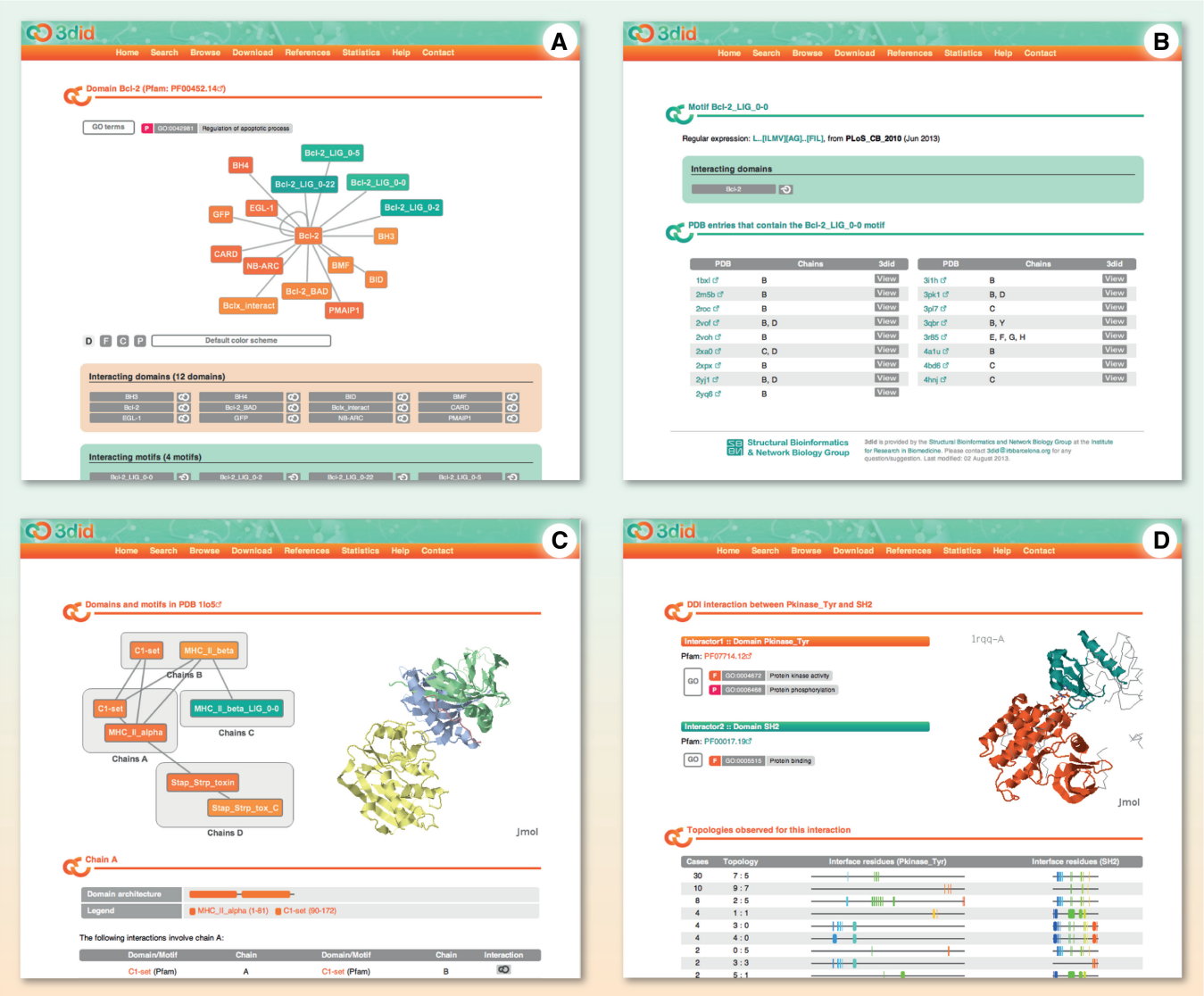

Figure 3. Views available in 3did. 3did provides four views to browse the data contained in the database: the Domain view (**A**), the Motif view (**B**), the PDB view (**C**) and the Interaction view (**D**).

DMIs. The second part of the page displays the residues that are involved in the interactions and the third one lists the structures in which the interactions have been identified and the chains that are involved. The second view, the motif view, lists the interacting domains and the structures in which the interactions have been identified. The third view, the interaction view, displays in a Jmol applet (http://www.jmol.org/) the first structure (in alphabetical order) in which the interaction has been identified. Different interfaces, involving different residues, may be identified for each interaction. Those different interfaces, or topologies, are also listed in the interaction page, with the residues involved. In addition, the list of structures in which the interaction has been found is provided: each structure can be displayed in Jmol by clicking the associated 'View' button. Finally, the PDB view allows displaying all domains and motifs identified in a specific PDB structure as well as their interactions. The DDIs and DMIs are represented as a network in CytoscapeWeb and the structure is displayed in Jmol. In addition, the domain architecture of each chain in the PDB file is listed, as well as the interactions in which each domain of the respective chain is involved (Figure 4B).

The navigation from one view to the others is facilitated by a number of links, including the clickable domains and motifs names, the DDI and domain–motif interaction buttons, the 'View' and 'Jmol' buttons, and the nodes and edges in the networks. The help page contains an illustrated description and additional information on how to browse the 3did web site.

CONCLUDING REMARKS

Full atomic characterization of protein–protein interaction at the 'omics level is becoming an impending need

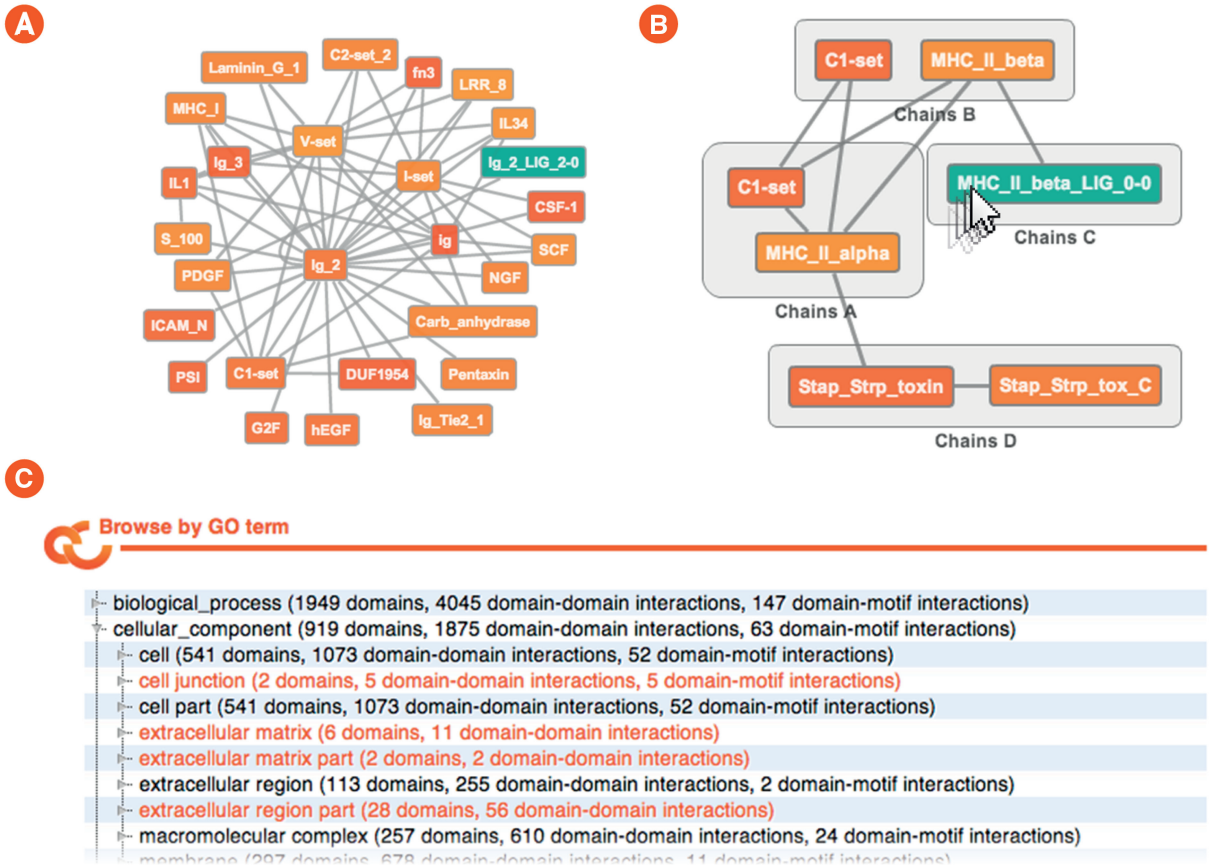

Figure 4. Browsing 3did. (**A**) Interactive view of the DDI and DMI network involving a particular domain. In orange are the domains while the motifs are in green. By clicking on any node or edge you are redirected to the page showing details about the corresponding domain, motif or interaction. (**B**) Interactive view of the domains and motifs in a PDB file. It shows the domain composition of the different chains (clustered on identical domain composition) as well as the motifs present in the chains. Lines connect domains and motifs that are interacting. Both nodes and lines can be clicked in order to visualize the details of the corresponding domain, motif or interaction. The CytoscapeWeb-based network visualizations (A and B) require a Flash plug-in to be installed in the browser to run. (**C**) Browse by GO term. A new tree view in the 'Browse' page allows searching for all the domains that are annotated with a specific GO-term.

in the everyday work of biologists (36). Many different approaches have been taken in order to achieve this. Most of them exploit the observation that evolutionary conserved domain families are used as independently interacting modules in proteins. These functional modules are reflected at the protein structural level and are involved in a complex network of interactions for which high-resolution structures are available in the PDB. 3did collects and organizes the catalog of these structures both for DDIs and DMIs. Furthermore, it makes the catalog available to the scientific community through an intuitive web interface for browsing the data and through batch downloads that enable the use of the data in large-scale bioinformatics studies. By providing a constantly updated, extensive catalog of 3D structures of domain-based interactions, 3did aims to be a reference resource for the structural annotation of protein interaction networks.

AVAILABILITY

3did can be accessed interactively from the web pages at http://3did.irbbarcelona.org, where it is also possible to download the full dataset in tab delimited files or in a full mysql dump that can be restored locally.

Four tab delimited files are available: 3did_flat.gz contains interacting domain pairs and the instances of these interactions in PDB structures, 3did_dmi_flat.gz contains DMIs, i.e. motifs with the corresponding pattern as well as all 3D instances of the interaction, 3did_interface_flat.gz contains the different binding topologies and 3did_global_interface_flat.gz contains the global interfaces.

More information about the download formats is available in the download page. 3did, including both DDIs and DMIs, will be updated twice per year with the latest versions of PDB and Pfam.

FUNDING

Funding for open access charge: The Spanish Ministerio de Ciencia e Innovación through the grant [BIO2010-22073] and the European Commission under FP7 Grant Agreement [306240 (SyStemAge)].

Conflict of interest statement. None declared.

REFERENCES

1. Yu,H., Braun,P., Yildirim,M.A., Lemmens,I., Venkatesan,K., Sahalie,J., Hirozane-Kishikawa,T., Gebreab,F., Li,N., Simonis,N. *et al.* (2008) High-quality binary protein interaction map of the yeast interactome network. *Science*, **322**, 104–110.
2. Arabidopsis Interactome Mapping Consortium. (2011) Evidence for network evolution in an Arabidopsis interactome map. *Science*, **333**, 601–607.
3. Rual,J.F., Venkatesan,K., Hao,T., Hirozane-Kishikawa,T., Dricot,A., Li,N., Berriz,G.F., Gibbons,F.D., Dreze,M., Ayivi-Guedehoussou,N. *et al.* (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature*, **437**, 1173–1178.
4. Rose,P.W., Bi,C., Bluhm,W.F., Christie,C.H., Dimitropoulos,D., Dutta,S., Green,R.K., Goodsell,D.S., Prlic,A., Quesada,M. *et al.* (2013) The RCSB Protein Data Bank: new resources for research and education. *Nucleic Acids Res.*, **41**, D475–D482.
5. Schuster-Bockler,B. and Bateman,A. (2007) Reuse of structural domain-domain interactions in protein networks. *BMC Bioinformatics*, **8**, 259.
6. Mosca,R., Ceol,A. and Aloy,P. (2013) Interactome3D: adding structural details to protein networks. *Nat. Methods*, **10**, 47–53.
7. Meyer,M.J., Das,J., Wang,X. and Yu,H. (2013) INstruct: a database of high-quality 3D structurally resolved protein interactome networks. *Bioinformatics*, **29**, 1577–1579.
8. Shoemaker,B.A., Panchenko,A.R. and Bryant,S.H. (2006) Finding biologically relevant protein domain interactions: conserved binding mode analysis. *Protein Sci.*, **15**, 352–361.
9. Davis,F.P. and Sali,A. (2005) PIBASE: a comprehensive database of structurally defined protein interfaces. *Bioinformatics*, **21**, 1901–1907.
10. Finn,R.D., Marshall,M. and Bateman,A. (2005) iPfam: visualization of protein-protein interactions in PDB at domain and amino acid resolutions. *Bioinformatics*, **21**, 410–412.
11. Gong,S., Yoon,G., Jang,I., Bolser,D., Dafas,P., Schroeder,M., Choi,H., Cho,Y., Han,K., Lee,S. *et al.* (2005) PSIbase: a database of Protein Structural Interactome map (PSIMAP). *Bioinformatics*, **21**, 2541–2543.
12. Yellaboina,S., Tasneem,A., Zaykin,D.V., Raghavachari,B. and Jothi,R. (2011) DOMINE: a comprehensive collection of known and predicted domain-domain interactions. *Nucleic Acids Res.*, **39**, D730–D735.
13. Jefferson,E.R., Walsh,T.P., Roberts,T.J. and Barton,G.J. (2007) SNAPPI-DB: a database and API of Structures, iNterfaces and Alignments for Protein-Protein Interactions. *Nucleic Acids Res.*, **35**, D580–D589.
14. Encinar,J.A., Fernandez-Ballester,G., Sanchez,I.E., Hurtado-Gomez,E., Stricher,F., Beltrao,P. and Serrano,L. (2009) ADAN: a database for prediction of protein-protein interaction of modular domains mediated by linear motifs. *Bioinformatics*, **25**, 2418–2424.
15. Pugalenthi,G., Bhaduri,A. and Sowdhamini,R. (2006) iMOTdb–a comprehensive collection of spatially interacting motifs in proteins. *Nucleic Acids Res.*, **34**, D285–D286.
16. Vanhee,P., Reumers,J., Stricher,F., Baeten,L., Serrano,L., Schymkowitz,J. and Rousseau,F. (2010) PepX: a structural database of non-redundant protein-peptide complexes. *Nucleic Acids Res.*, **38**, D545–D551.
17. Das,A.A., Sharma,O.P., Kumar,M.S., Krishna,R. and Mathur,P.P. (2013) PepBind: a comprehensive database and computational tool for analysis of protein-peptide interactions. *Genomics Proteomics Bioinformatics*, **11**, 241–246.
18. Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
19. Sillitoe,I., Cuff,A.L., Dessailly,B.H., Dawson,N.L., Furnham,N., Lee,D., Lees,J.G., Lewis,T.E., Studer,R.A., Rentzsch,R. *et al.* (2013) New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. *Nucleic Acids Res.*, **41**, D490–D498.
20. Punta,M., Coggill,P.C., Eberhardt,R.Y., Mistry,J., Tate,J., Boursnell,C., Pang,N., Forslund,K., Ceric,G., Clements,J. *et al.* (2012) The Pfam protein families database. *Nucleic Acids Res.*, **40**, D290–D301.
21. Stein,A., Russell,R.B. and Aloy,P. (2005) 3did: interacting protein domains of known three-dimensional structure. *Nucleic Acids Res.*, **33**, D413–D417.
22. Stein,A., Panjkovich,A. and Aloy,P. (2009) 3did Update: domain-domain and peptide-mediated interactions of known 3D structure. *Nucleic Acids Res.*, **37**, D300–D304.
23. Stein,A., Ceol,A. and Aloy,P. (2011) 3did: identification and classification of domain-based interactions of known three-dimensional structure. *Nucleic Acids Res.*, **39**, D718–D723.
24. Chakrabarti,P. and Janin,J. (2002) Dissecting protein-protein recognition sites. *Proteins*, **47**, 334–343.
25. Finn,R.D., Clements,J. and Eddy,S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.*, **39**, W29–W37.
26. Aloy,P. and Russell,R.B. (2002) Interrogating protein interaction networks through structural biology. *Proc. Natl Acad. Sci. USA*, **99**, 5896–5901.
27. Aloy,P. and Russell,R.B. (2003) InterPreTS: protein interaction prediction through tertiary structure. *Bioinformatics*, **19**, 161–162.
28. Scott,J.D. and Pawson,T. (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science*, **326**, 1220–1224.
29. Ejiri,K., Taniguchi,H., Ishihara,K., Hara,Y. and Baba,S. (1990) Possible involvement of cholinergic nicotinic receptor in insulin release from isolated rat islets. *Diabetes Res. Clin. Pract.*, **8**, 193–199.
30. Dinkel,H., Michael,S., Weatheritt,R.J., Davey,N.E., Van Roey,K., Altenberg,B., Toedt,G., Uyar,B., Seiler,M., Budd,A. *et al.* (2012) ELM–the database of eukaryotic linear motifs. *Nucleic Acids Res.*, **40**, D242–D251.
31. Edwards,R.J., Davey,N.E., O'Brien,K. and Shields,D.C. (2012) Interactome-wide prediction of short, disordered protein interaction motifs in humans. *Mol. Biosyst.*, **8**, 282–295.
32. Hugo,W., Song,F., Aung,Z., Ng,S.K. and Sung,W.K. (2010) SLiM on Diet: finding short linear motifs on domain interaction interfaces in Protein Data Bank. *Bioinformatics*, **26**, 1036–1042.
33. Stein,A. and Aloy,P. (2010) Novel peptide-mediated interactions derived from high-resolution 3-dimensional structures. *PLoS Computat. Biol.*, **6**, e1000789.
34. The Gene Ontology Consortium. (2013) Gene Ontology annotations and resources. *Nucleic Acids Res.*, **41**, D530–D535.
35. Lopes,C.T., Franz,M., Kazi,F., Donaldson,S.L., Morris,Q. and Bader,G.D. (2010) Cytoscape Web: an interactive web-based network browser. *Bioinformatics*, **26**, 2347–2348.
36. Mosca,R., Pons,T., Ceol,A., Valencia,A. and Aloy,P. (2013) Towards a detailed atlas of protein-protein interactions. *Curr. Opin. Struct. Biol.*, **23**, 929–940.